ACCEPTED MANUSCRIPT

Towards Developing an Open-Infrastructure for Assessing the Progress, Success, and Impacts of CyberInfrastructure Projects

This Accepted Manuscript (AM) is a PDF file of the manuscript accepted for publication after peer review, when applicable, but does not reflect post-acceptance improvements, or any corrections. Use of this AM is subject to the publisher's embargo period and AM terms of use. Under no circumstances may this AM be shared or distributed under a Creative Commons or other form of open access license, nor may it be reformatted or enhanced, whether by the Author or third parties. By using this AM (for example, by accessing or downloading) you agree to abide by Springer Nature's terms of use for AM versions of subscription articles: https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

The Version of Record (VOR) of this article, as published and maintained by the publisher, is available online at: https://doi.org/10.1007/s42979-024-02961-8. The VOR is the version of the article after copy-editing and typesetting, and connected to open research data, open protocols, and open code where available. Any supplementary information can be found on the journal website, connected to the VOR.

For research integrity purposes it is best practice to cite the published Version of Record (VOR), where available (for example, see ICMJE's guidelines on overlapping publications). Where users do not have access to the VOR, any citation must clearly indicate that the reference is to an Accepted Manuscript (AM) version.

Accepted manuscript

ACCEPTED MANUSCRIPT

Towards Developing an Open-Infrastructure for Assessing the Progress, Success, and Impacts of CyberInfrastructure Projects

*Ritu Arora, Wayne State University, Michigan, USA (ritu@wayne.edu)

Sukrit Sondhi, MacMillan Learning, Texas, USA (sukrit@yahoo.com)

Abstract: **Purpose**: Metrics for objectively assessing the CyberInfrastructure (CI) projects are important for not only obtaining the required resources for the long-term sustainability of the projects and community engagement but also for identifying issues and prioritizing the areas of improvement. This paper presents an overview of a model named MICI for assessing the progress, success, and impacts of a diverse range of CI projects. **Methods**: As an initial step, we developed a taxonomy of the CI projects and defined a set of standard metrics for the assessment of the projects belonging to the different categories in the taxonomy. If all the projects in the same category adopt a standard set of metrics for their assessment *in addition to* their project-specific and general metrics for assessment, it can help in focusing on gathering the assessment related data uniformly and understanding the short-term and long-term impacts created by the investments and the projects. We also introduce the MICI model for measuring the impacts of CI projects. **Results**: Models like MICI along with the periodic peer-review can be useful in standardizing the process of assessing the impacts of the CI projects while ensuring accountability and transparency. **Conclusion**: *With community engagement,* the taxonomy of the CI projects, the associated metrics, and the impacts model should be refined further for ensuring their continued relevance and usefulness. Additionally, quantitative metrics should be used to supplement human judgment in the evaluation process and not to replace it.

Keywords: CyberInfrastructure, CI, metrics, taxonomy, impact, success, U.S. federal government funding

1. Introduction

CyberInfrastructure (CI) can be defined as the information technology infrastructure that offers advanced capabilities for supporting any or all stages of research and discovery. It consists of **hardware** such as High Performance Computing (HPC) platforms, cloud computing platforms, data storage systems, data repositories, visualization equipment, sensors, and edge-computing devices. It also consists of **high-speed networks** and the associated middleware. Additionally, a wide variety of **software and data** products are also a part of the CI. The **people and processes** that are involved in the design, development, deployment, operation, maintenance, and usage of the CI are considered as a part of the CI as well. Thus, there is a wide spectrum of CI projects in the community, ranging from provisioning hardware resources to developing software tools and services to workforce development. These projects **involve billions of dollars of investments from the U.S. federal funding agencies.** As an example, the National Science Foundation (NSF) has invested more than $650 million over a decade on a diverse range of CI software and data products through their programs such as Cyberinfrastructure for Sustained Scientific Innovation (CSSI) [1], Software Infrastructure for Sustained Innovation (SI2) [2], and Data Infrastructure Building Blocks (DIBBs) [3]. Given the scale of federal investments, it is important to have a mechanism for the transparent and periodic assessment of the Progress, Success, and Impacts (PSI) created by the CI projects and their funding programs. Such assessments can help in (1) identifying the issues, if any, with the projects, (2) prioritizing the areas of improvement, (3) ensuring accountability of spendings on the projects, (4) connecting the progress/outcome of the projects to societal/technology impacts, (5) ensuring the availability of the resources for their long-term sustainability, (6) measuring the Return On Investment (ROI) for informing the future directions of the investments, and (7) creating opportunities for community engagement. The **assessments should help in capturing metrics** such as the number of people in the

Accepted manuscript

ACCEPTED MANUSCRIPT

community that are benefitting from the investments and the ways in which the CI projects have advanced science and the prosperity of the nation.

In general, the impact of a project can be defined as the influence or effect that it creates on the Principal Investigator (PI), project team, PI's institution, scientific domain, economy, society, environment, technology, processes, and policies. It can include transformations or advances in the domains related to the project, creation of new knowledge, advancement of skills of the different stakeholders, improvement in the quality of education, increase in the performance and productivity of the people, and new opportunities for strengthening Diversity, Equity, Inclusion, and Accessibility (DEIA). For measuring the level of impact created by the projects, **it is important to define a set of well-thought, project-specific metrics**, and monitor how those metrics evolve or change with the progress of the projects throughout their lifecycles.

As the CI projects can vary widely in type, size, duration, goals, resources, and other factors, and each project could be having its own set of **project-specific metrics**, it can be challenging to conduct an assessment to understand the overall impact created by the investments and analyze patterns or trends over time**.** Additionally, there is no well-defined common framework for measuring CI project metrics and calibrating them. Due to this, different projects may be using different metrics to measure the same thing, thereby, making it difficult to compare data from different sources and develop field-wide benchmarks. Therefore, in addition to using the project-specific metrics, it is **important to define and adopt a standardized set of metrics and processes** that are applicable across a wide variety of CI projects. Having a metrics strategy that can be tailored to project-specific needs and goals while also incorporating common frameworks for measuring CI metrics will facilitate the comparison of data from different sources and the development of industry-wide benchmarks.

It is important that the standardization process such as mentioned above is (1) community-driven, (2) sensitive to the diversity of CI projects, disciplines and institution sizes, and (3) respects the autonomy or creative independence of the projects' PIs. The standardization process should have the potential of contributing towards maximizing the quality and impact of the CI projects and their associated funding programs. Additionally, the standardization process should be designed carefully such that it does not place any undue burden on the PIs. If adopted by the community, such a standardization process can help in not only uniformly measuring the PSIs of the CI projects but also their associated funding programs for the current and future decision-making purposes.

As a first step towards the standardization process, it is important to create **a taxonomy of the CI projects, and map each category in the taxonomy to a standard set of metrics for assessing their short-term and long-term impacts.** The short-term impacts are created during the lifecycle of the project or shortly after it is over. The long-term impacts are related to the sustainable use and impact created by the projects even after they are over years ago. During the creation of the taxonomy of CI products/projects, it is also important to: create dictionaries so that there is a common understanding of the elements used in the taxonomy, define approaches or recipes for data collection and for using the collected data to measure impact, create an open data infrastructure that supports capturing information on both quantitative and qualitative metrics for assessing the impact created by the CI projects, provide methods to derive insights from the collected data, and create a unique identification number for each CI product/project and make them findable by listing them in a catalog. The data collected through standard metrics can be used as an input to a model (or a framework) that can then help in understanding the overall impact created by the CI projects and the funding programs. Such a model should be designed such that quantitative metrics or indicators are used responsibly and manipulation of the metrics is discouraged. In the pursuit of defining such a standardization process, this paper makes the following contributions:

1. Introduces a preliminary taxonomy of the CI projects,

Accepted manuscript

ACCEPTED MANUSCRIPT

2. Includes an overview of the process for standardization of the metrics and introduces a metrics model for CI projects (named as MICI),
3. Provides examples of CI projects and demonstrates the application of the MICI model, and
4. Discusses the prototype of an open-infrastructure for capturing metrics (Opuntia project).

In the rest of this paper, the taxonomy of CI projects is introduced in Section 4 and a set of standard metrics for evaluating the projects in each category are described in Sections 5 and 6. We discuss a model for assessing the impacts of the projects in Section 7. We then discuss an open-infrastructure for gathering metrics in Section 8 of the paper.

2. Background and Related Work

While the focus of this paper is mainly on CI projects, there is prior related work in the areas of measuring impacts of research management, research infrastructure, and healthcare. In this section, we present some of the assessment frameworks that are most closely related to our work. We also explain the topic of "responsible metrics".

2.1. DORA Quality of Assessment Framework

DORA stands for the Declaration On Research Assessment [4]. The DORA Quality of Assessment Framework (QAF) is a movement to improve the ways in which the output of scholarly research is evaluated. The QAF is a set of guidelines for assessing the quality of research projects and their outputs, and it is intended to be used by a variety of stakeholders, including funding agencies, universities, and researchers themselves. The QAF provides a set of criteria for assessing the quality of research projects and their outputs. These criteria are based on the following dimensions:

- **Research design and planning:** This dimension assesses the quality of the research question, the research methods, and the data collection and analysis plan.
- **Research execution:** This dimension assesses the quality of the data collection and analysis, and the soundness of the conclusions.
- **Research dissemination and impact:** This dimension assesses the quality of the dissemination of the research findings, and the impact of the research on the field.

The QAF also emphasizes the importance of ethical considerations in research assessment. The framework states that "*research should be conducted in an ethical manner, and that research assessment processes should also be ethical*."

The QAF does not endorse any specific metrics for assessing research quality. However, the framework does acknowledge that metrics can be useful tools for assessing quality, if used carefully and in conjunction with other information.

The QAF states that "*metrics can be used to provide quantitative information about research projects and their outputs. However, **it is important to note that metrics do not provide a complete picture of research quality.** It is important to consider a variety of factors when assessing research quality, including the metrics, but also the other dimensions of quality described in this framework.*"

The QAF also provides some **guidance on how to use metrics responsibly (explained further in Section 2.5)**. The framework states that "when using metrics to assess research quality, it is important to consider the following:

- The relevance of the metric to the research question and the field of study
- The limitations of the metric
- The context in which the metric is being used
- The potential for bias in the metric

Accepted manuscript

ACCEPTED MANUSCRIPT

Overall, the QAF provides a valuable framework for thinking about research quality assessment. The framework emphasizes the importance of considering a variety of factors, including metrics, when assessing research quality.

2.2. UK Research Excellence Framework (REF)

The Research Excellence Framework (REF) [5] is a UK-wide system for assessing the quality of research in Higher Education Institutions (HEIs). The REF is carried out every seven years, and it is used to inform the allocation of research funding to HEIs.

The REF is based on the following principles:

- Research quality is **assessed by peers**: The REF is assessed by panels of experts in each field of research.
- Research quality is **assessed in context**: The REF takes into account the different contexts in which research is conducted, such as the size and resources of the HEI.
- Research quality is **assessed based on a variety of factors**: The REF considers a variety of factors, such as the significance of the research, the rigor of the methodology, and the impact of the research.

The REF is assessed in three areas that are outlined below and also shown in Figure 1:

- **Outputs:** The REF assesses the quality of research outputs, such as papers published in peer-reviewed journals, books, and patents.
- **Impact:** The REF assesses the impact of research on society, the economy, and culture.
- **Environment:** The REF assesses the research environment at each HEI, including the quality of research staff, facilities, and infrastructure.

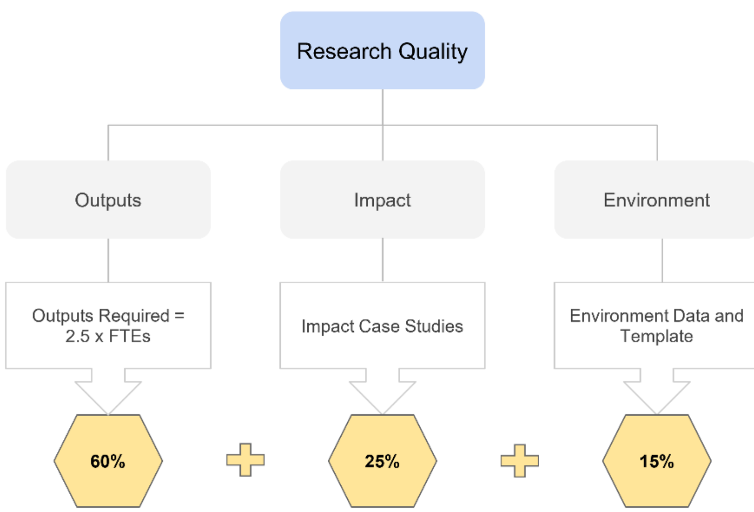


Figure 1. Overview of REF assessment areas [6]

REF does not rely on quantitative metrics alone and assesses the quality of output according to originality, significance, and rigor. Likewise, the quality of impacts is assessed according to reach, and significance. The environment quality is assessed according to vitality and sustainability.

ACCEPTED MANUSCRIPT

2.3. Snowball Metrics Framework

The Snowball Metrics framework [7] is a set of metrics that can be used to measure the performance and impact of research projects, and benchmark institutions of higher education. The framework is based on the idea that research is a cumulative process, and that the success of a research project is often dependent on the success of previous research projects. Snowball Metrics have been defined for the entire spectrum of research activities, such as, funding, collaboration, publication, and commercialization.

The Snowball Metrics framework is designed to be flexible and adaptable, so the specific metrics that are used will vary depending on the nature of the research project. However, the following three categories of metrics can be used as a starting point for developing a metrics framework for any research project: **research inputs, research processes, and research outputs.**

At the time of writing this paper, there are 32 recipes or methodologies available for calculating the Snowball Metrics in a consistent way for inputs, processes, and outputs or outcomes related to each of the following sectors in the research landscape: research unit, enterprise activities/economic development, and postgraduate education. These are shown in Table 1. As an example, let us consider inputs, processes, and outputs/outcomes metrics for a research enterprise.

Research inputs: With the metrics for research input, a research unit at a university can track the volume of applications for funding that were submitted, number of funding applications that were approved, and the success rate.

Research processes: The metrics for the research process can include the volume of research expenditure at an organization and can be used to measure the increase or decrease in the work undertaken on external funding. It can also include the market share.

Research outputs and outcomes: The examples of metrics for assessing research outputs and outcomes can include the citation count, academic-industry collaborations, altmetrics, public engagement, and academic recognition.

Table 1 shows the 32 Snowball Metrics and the explanation on these is available in the Snowball Metrics recipe book [7]. Apart from deciding on the set of metrics, the stakeholders also decide on a set of denominators. The denominators are needed to (1) "Slice and dice" the Snowball Metrics at different levels of granularity, such as, at the entire institution level, department level, and discipline level, and (2) normalize the data to account for the differences in the sizes of institutions to ensure that larger institutions are not always appearing to perform better than the smaller ones.

Table 1. Snowball Metrics

	Research Inputs	**Research Processes**	**Research Outputs and Outcomes**
Research	Application Volume, Awards Volume, Success Rate	Income Volume, Market Share	**Publications & Citations** Scholarly Output, Citation Count, Citations Per Output, h-Index, Field-Weighted Citation Impact, Outputs in Top Percentiles, Publications in Top Journal Percentiles **Collaboration** Collaboration, Collaboration

Accepted manuscript

ACCEPTED MANUSCRIPT

			Publication Share, Collaboration Impact, Collaboration Field-Weighted Citation Impact, Academic-Corporate Collaboration, Academic-Corporate Collaboration Impact **Societal Impact** Altmetrics, Public Engagement, Academic Recognition
Enterprise Activities/Economic Development	Academic-Industry Leverage, Business Consulting Activities	Contract Research Volume	Intellectual Property Volume, Intellectual Property Income, Sustainable Spin-Offs, Sustainable Spin-Off Related Finances
Post-Graduate Education	Research Student Funding	Research Student to Academic Staff Ratio	Time to Award of Doctoral Degree, Destination of Research Student Leavers

2.4. STAR Metrics

The Science and Technology for America's Reinvestment (STAR) metrics program [8] was a joint effort involving the National Institutes of Health (NIH), the National Science Foundation (NSF), the White House Office of Science and Technology Policy (OSTP), the U.S. Department of Agriculture (USDA), and U.S. Environment protection Agency (EPA), with the objective of documenting the ROI, research impact, and outcomes of national investments in research. Initially, a limited number of U.S. universities contributed a limited number of data fields from their existing university administrative databases to report on the impact of federal research and development spending on job creation. Later several other organizations including DOE joined the effort and contributed the data in a central database named Federal RePORTER. This database also included information that was automatically collected from other sources such as Medline and PubMed Central. The STAR Metrics project was discontinued in 2022 and recent information on its current status is unavailable to us.

2.5. Responsible Metrics

The impact of research work in academia is often measured using a variety of metrics, including:

- **Citation counts:** The number of times a research publication has been cited by other researchers is a common measure of its impact. Citation counts can be measured using a variety of databases, such as Google Scholar, Web of Science, and Scopus.

- **Journal impact factor (JIF):** The JIF is a measure of the average number of citations received by articles published in a particular journal over a two-year period. It is often used to assess the quality of a journal and the impact of the research published in it.

Accepted manuscript

ACCEPTED MANUSCRIPT

- **H-index:** The h-index is a measure of an individual researcher's productivity and impact. It is calculated by counting the number of papers a researcher has published that have each been cited at least h times.

- **Altmetrics:** Altmetrics are alternative metrics of research impact that measure the online attention and engagement that research publications receive. They can include metrics such as social media shares, downloads, and mentions in the news.

However, these metrics also have some inherent drawbacks and limitations:

- **Citation counts** can be inflated by self-citation. Groups of researchers can agree to cite each other's work excessively. Citation counts can also be biased towards certain types of research, such as quantitative research published in high-impact journals. Plus, citation counts can take time to accumulate, so they may not be a good measure of the immediate impact of research.
- **Journal impact factors (JIF)** are biased towards certain types of journals, such as general-interest journals that publish a wide range of research, and also do not take into account the quality of individual articles, only the average number of citations to articles published in the journal.
- **The H-index** is also biased towards researchers who have published a large number of papers in high-impact journals and does not take into account the quality of individual articles, only the number of citations to the researcher's work.
- **Altmetrics** can be used to measure the reach and engagement of research, but they are not a good measure of the quality or impact of research. They can be biased towards certain types of research, such as research that is controversial or that has a high public interest.

The Coalition of Advanced Research Assessment (CoARA) agreement [9] for reforming research assessment states the following commitments for its signatories that are directly relevant to the topic of this paper: (1) “Abandon inappropriate uses in research assessment of journal- and publication based metrics,” and (2) “Base research assessment primarily on qualitative evaluation for which peer review is central, supported by **responsible use** of quantitative indicators”.

The DORA declaration [4] states that certain metrics even though commonly used (such as, the journal impact factor) as a surrogate for the quality of research are not a true quality measure and their use in assessments should be discouraged. In a 2015 report [10] titled “The Metric Tide” that was commissioned by the U.K. Higher Education Funding Council, the authors present a set of “responsible metrics” for research assessment and management. They share five principles underpinning “responsible metrics” and these include: (1) **robustness** - so that the metrics are based on the accurate data that is in the scope of the assessment, (2) **humility** - to recognize that quantitative assessment should supplement the peer-reviews and not replace them, (3) **transparency** - in keeping data collection and analytical process open for evaluation and verification, (4) **diversity** - in the range of metrics and accounting for the variations by fields, (5) **reflexivity** - to recognize and anticipate the effects of indicators/metrics and updating them as needed.

3. Responsible Metrics Use for CI

We analyzed the use of “responsible metrics” in the context of CI projects. We noticed that certain metrics are commonly used for assessing the CI projects but they may not convey the true impact created by those projects. For example, the number of downloads of a software is used as a metric, however, the number of downloads alone does not indicate if a software was actually used and achieved its intended impact. Additionally, multiple downloads of a given software could be done by a single user or a bot, thereby inflating the number of total downloads for a software without actually using the software. In this situation, perhaps capturing the information on the IP addresses and geographical locations from where the downloads were triggered, along with the email addresses of those downloading the software, could provide

Accepted manuscript

ACCEPTED MANUSCRIPT

a mechanism for follow-ups and gathering further qualitative information on software use and its impacts. This example highlights the **need for clearly defining what are acceptable metrics for assessing the impacts of CI projects and what are the methods to capture and report those metrics in an ethical manner**. Engagement of the community and funding agencies is needed though to fulfill this need in a sustainable manner.

Quantitative metrics are those that have numerical values, and these values can be measured directly or derived using mathematical formulae. In contrast, **qualitative metrics** are those that are subjective - they are based on human judgment and/or descriptive information. **Human judgment in the form of expert opinion (or qualitative metrics) is very important for establishing the connection between the projects and their claimed impacts.** Therefore, we have adopted a mix of both quantitative and qualitative metrics in selecting the list of metrics for assessing CI projects and measuring their impacts.

As shared in Section 6 of the paper, we have curated lists of metrics for the different categories of the CI projects. These metrics are in line with the principles of CoARA and DORA. We have incorporated the concept of responsible use of metrics and have been mindful about not including the metrics such as the journal quality in the assessment of project outcomes. In addition to defining the category-specific metrics (in Section 6), it is also important to abstract the general metrics for measuring the progress, success, and impact of the projects as described in Section 5.

4. Classification of CI Projects

One of the key challenges to the usefulness of metrics is the applicability and availability of a common set of metrics across a diverse set of projects. In the case of CI projects, the unique nature of each project makes it difficult to apply the same yardstick (metrics) across projects. CI projects can have very different focus areas, such as projects for the design, development, and deployment of new CI resources, integration of existing CI resources, provisioning and support services for CI resources, development and delivery of CI training and education, and process improvement, security and compliance.

To manage the diversity and complexity in the range of the CI projects and to do a fair comparison, it is important to classify the projects into categories according to a standard taxonomy. Next, a standard set of metrics should be defined for each project category. The comparison of category-level metrics associated with different projects will likely enhance our knowledge and decision-making.

As a part of an NSF-funded project named as Opuntia [11] we have developed a **taxonomy of CI projects** which is detailed in “Appendix A: A Taxonomy for CI Projects” and its **high-level categories are as follows**: (1) Hardware or instruments, (2) Software, (3) Data, (4) Processes, (4) People, (5) Combination of multiple categories, and (6) Other. The category of “Software” has a subcategory of “Data Management” which covers software related to areas such as “Data Privacy”, “Data Protection and Recovery” and “Semantic Web”. Likewise, other categories in the taxonomy will be refined in the future with community engagement and subcategories will be added as applicable. In its current form though, this taxonomy is sufficient as a starting point for the illustration of the concept of standardization of metrics presented in this paper.

5. General Metrics for Progress, Success, and Impact of CI Projects

The kind of metrics associated with a project can vary over the course of the project’s life cycle. Depending upon whether the project is in-progress or complete, the metrics can be divided into three dimensions - Progress, Success, and Impact (**PSI**) - and Figure 2 shows an overview of these dimensions and some of the metrics that are related to each of the dimensions. These metrics can be applicable to any project and are not specific to CI projects.

Accepted manuscript

5.1. Progress

The progress on task completion or for achieving the project goals are measured under this dimension. These can be further understood as percentage of completion of tasks, number of milestones achieved, number of risks mitigated, and the type of feedback received.

- **Percentage complete by tasks, effort, time, and cost:** These metrics can be used to track the project's progress against its original goals and objectives. The **percentage completion by tasks** can be calculated by dividing the number of tasks that have been completed by the total number of tasks in the project. The **percentage completion by effort** can be calculated by dividing the amount of effort that has been expended on the project by the total amount of effort that was estimated for the project. The **percentage completion by time** can be calculated by dividing the amount of time that has elapsed since the project began by the total amount of time that was estimated for the project. The **percentage completion by cost** can be calculated by dividing the amount of money that has been spent on the project by the total amount of money that was budgeted for the project.

- **Milestones achieved:** Milestones are key events in the project lifecycle. They mark the completion of a significant phase of the project or the delivery of a major deliverable. Tracking progress towards milestones can help to identify potential delays and ensure that the project is on track.

- **Risks mitigated:** Risks are events or conditions that could have a negative impact on the project. Risk mitigation is the process of identifying and taking steps to reduce the risk of these events or conditions occurring. Tracking the number of risks that have been mitigated can help to assess the overall risk profile of the project.

- **Positive feedback from periodic reviews:** Periodic reviews are held to assess the progress of the project and to identify any areas where improvement is needed. Positive feedback from periodic reviews can be a good indication that the project is on track and that it is meeting the expectations of stakeholders.

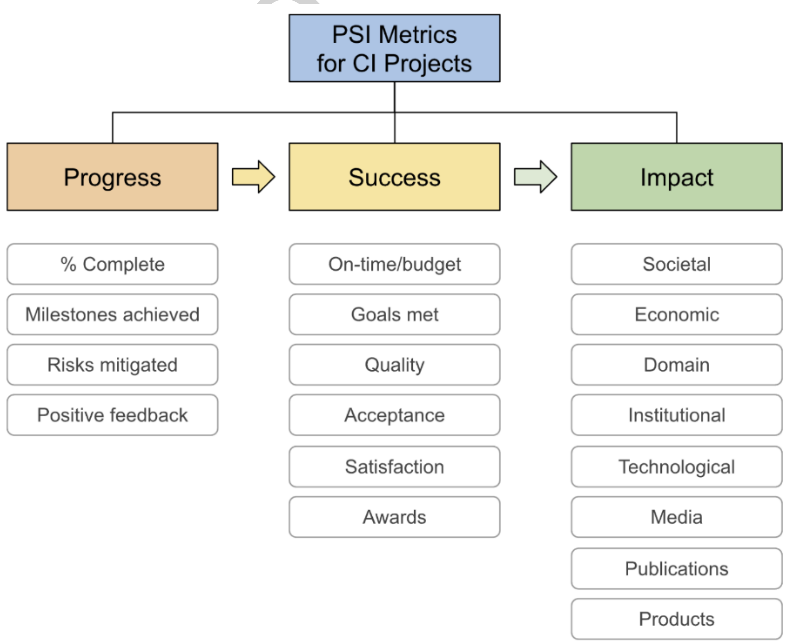


Figure 2. Overview of PSI Metrics for projects

Accepted manuscript

In this paper, we do not attempt to define the criteria for "project end", as that can vary according to the circumstances of different projects and their sponsors/governors. However, some of the common criteria under which projects are considered to have ended are:

- The funding has run out.
- Work has stopped due to any reason and is not expected to resume.
- Project goals have been achieved or are now considered unachievable.
- The project was cancelled before the planned completion due to any reason.

5.2. Success

The **required output** for meeting the project goals and making the projects **successful** are measured under this dimension. The expected output and the definition of success will likely vary from one CI project to another. However, mentioned below are some general categories of metrics for assessing the success of the projects.

- **On-time and on-budget delivery:** This metric measures whether the project was completed on time and within budget. This is an important metric because it indicates whether the project was well-managed and whether the project team was able to deliver on its commitments.

- **Outcomes - meeting project goals:** This metric measures whether the project met its stated goals and objectives. This is the most important metric because it indicates whether the project was successful in achieving its intended purpose.

- **Quality - meeting or exceeding project requirements:** This metric measures whether the project deliverables met or exceeded the stated requirements. This is an important metric because it indicates whether the project deliverables are of high-quality and whether they will be useful to users.

- **Acceptance/adoption by customers or users:** This metric measures whether the project deliverables were accepted and adopted by customers or users. This is an important metric because it indicates whether the project deliverables are meeting the needs of users and whether they are having a positive impact.

- **Satisfaction - positive feedback from users and stakeholders:** This metric measures the level of satisfaction of users and stakeholders with the project deliverables. This is an important metric because it indicates whether the project is meeting the expectations of its stakeholders.

- **Any awards (such as Nobel Prizes or Turning Awards) won:** This metric measures whether the project won any awards, such as Nobel Prizes or Turning Awards. This is a rare but important metric because it indicates that the project has made a significant contribution to the field of science or technology.

5.3. Impact

The impact created by the projects is measured under this dimension in various contexts such as, society, economy, scientific domain, institution, technology, community, collaborations, and academic recognition. These are further discussed below.

- **Societal impact - such as policy changes:** This metric measures the impact of a project on society as a whole. This can include changes to government policies, regulations, or programs. For example, a project that develops a new tool for predicting natural disasters could lead to changes in disaster preparedness and response policies.

Accepted manuscript

ACCEPTED MANUSCRIPT

- **Economic impact - such as jobs created and businesses supported:** This metric measures the economic impact of a project. This can include the number of jobs created, the number of businesses supported, the number of revenue generating businesses or services started as a result of the project, and the amount of revenue generated. For example, a project that develops a new platform for online education could create jobs in the education and technology sectors.

- **Domain impact - such as advancing a field of research:** This metric measures the impact of a project on one or more domains. This can include the number of new scientific discoveries made, the number of new patents awarded, the income generated from the intellectual property, and the number of new products or services developed. For example, a project that develops a new HPC platform could advance the field of computational physics.

- **Institutional impact - such as, enhancing resources, reputation, skills, and capabilities:** This metric measures the impact of a project on the institution that received the funding. This can include the acquisition of new resources, such as equipment or funding, the enhancement of the institution's reputation, and the development of new skills and capabilities among the institution's staff. For example, a project that develops a new research data center is likely to enhance the institution's ability to support research and education.

- **Technological advancement:** This metric measures the technological advancement achieved by a project. This can include the development of new algorithms, software, hardware, or other technologies. For example, a project that develops a new machine learning algorithm could advance the field of artificial intelligence.

- **Media coverage:** This metric measures the amount of media coverage received by the cyberinfrastructure project. This can indicate the level of public interest in the project and its potential impact. For example, a project that is featured in a major news publication has the potential of generating public interest and support for the project.

- **Publications based on the project:** This metric measures the number of publications based on the project. This can indicate the impact of the project on the academic community. For example, a project that helps in generating new scientific results could lead to the publication of new research papers.

- **New products, patents, or services:** This metric measures the number of new products or services developed in a project. This can indicate the economic impact of the project and its potential to benefit society. For example, a project that is used to develop a new drug could lead to the development of a new product that benefits public health.

- **Collaborations:** The number of new collaborations fostered by a project locally, regionally, nationally, and globally, can be critical for community development and long-term sustainability of the projects. The partnerships formed with the industry can lead to taking the project output to market and creating economic impact. As an example, a company manufacturing a new processor can contribute towards extending a popular code profiling tool developed in a project to function with their new processor.

- **Community engagement**: Surveys or interviews for capturing the feedback/testimonials from the target audience of a project can help in assessing the project's impact on the community. Additionally, the number of events or activities organized or attended for community-building, marketing, and public dissemination of the results can be captured. **Altmetrics** can also be used to measure the reach and engagement of research without measuring the quality of the impact though.

Accepted manuscript

ACCEPTED MANUSCRIPT

- **Academic recognition**: The impact of the projects on the advancement of the PIs' career can also be considered for assessment of academic recognition. To measure this impact systematically though, a standard mechanism is needed for (1) mapping the different job profiles or titles at the PIs' institutions to a standard set of job categories, and (2) assigning a unique identifier to the products resulting from the CI projects, and (3) gathering all archivable assets and metrics related to the CI project in a central repository. If the PIs share the information about the career growth enabled by a CI project by associating the unique product numbers then that can help in collecting and aggregating data as needed to measure the impact of a CI project on PIs' career.
- **Educational impact**: This metric can help in capturing the improvements to the quality of education by measuring the number of courses created or enhanced as a result of a project and the number of theses completed.

6. Project Category-Specific Metrics

The three dimensions of metrics mentioned earlier - Progress, Success, and Impact (PSI) - lead to having different types of metrics for each of the project categories defined in the taxonomy. Some of the metrics for each dimension of the projects in different categories are shared in Tables 2 to Table 6. In the case of multidisciplinary projects, the PSI metrics defined under multiple categories can be used and additional metrics, such as, the number of training sessions on team science, can be gathered. We have likely not covered all the categories of projects and their relevant metrics at the time of writing this paper and hence, in future, additional categories of projects can be added to the taxonomy and their PSI metrics can be defined to support evolution in the CI landscape in a flexible manner. Additionally, **the lists of metrics mentioned in these tables are not exhaustive but are representative samples of the type of metrics that can be used for those categories.**

Table 2. PSI and Other Metrics for Hardware Projects

Progress	**Success**	**Impact**	**Other**
% of required components procured % of required components installed % of required components tested or benchmarked % of required components put into production Time taken for deployment Mean Time Between Failure (MTBF)	% of hardware components that meet or exceed requirements % of project goals achieved by the hardware Variation from project schedule and budget Reliability and availability % of total number of compute jobs submitted that ran successfully	Number of adopters or users from diverse domains Number of novel projects, products, or breakthroughs enabled Reduction in time-to-solutions Reduction in carbon footprint on the environment Increase in energy efficiency Increase in user productivity by decrease in	Heat dissipation: measure of the cooling requirements of the hardware to prevent overheating Network bandwidth: measure of the performance of the network and identification of bottlenecks Storage capacity: measure of the amount of data that hardware can store Compute power: measure of the processing power of the hardware

Accepted manuscript

ACCEPTED MANUSCRIPT

Mean Time To Repair (MTTR) % completion of the development of user guide and training content	System response time or latency Rate of system utilization	effort in porting applications Global recognition for the organization, nation and hardware, by featuring in the Top500 list	Performance per watt per dollar Quality of Service (QoS) - queue wait time & wall-time for job completion

Table 3. PSI Metrics for Software Projects

Progress	**Success**	**Impact**
% of completed software modules Time taken to develop and deploy new software features Code coverage - amount of the software code that has been tested Number of software modules tested that are working properly Number of code commits or changes or bug fixes in the code repository % completion of development of user guide, training content, or education and engagement activities for increasing software adoption User satisfaction with software features and functionality MTBF causing interruption in software usage and MTTR required for fixing errors in the software The number of iterations of software development Rate of integration and deployment	Software meets or exceeds the given requirements Software is deployed on time and within budget Software is reliable and available to users Software is usable A reasonable size of user community exists Performance and responsiveness of the software meets or exceeds the expectations Software scalability - software can handle increased load Software supports the targeted business functions Software is secure - there are no known vulnerabilities	Increase in number of adopters or users from diverse domains Number of novel projects, products, or breakthroughs enabled Digital accessibility of the software Degree of code reusability (1-5 rating) Degree of code interoperability (1-5 rating) Degree of software sustainability (1-5 rating) Number of contributors

Table 4. PSI Metrics for Data Projects

Progress	**Success**	**Impact**

Accepted manuscript

ACCEPTED MANUSCRIPT

Volume of data collected Volume of data cleaned Volume of data processed Volume of data made available to users	Data meets all requirements, including quality (accuracy, completeness, and consistency), format, and compliances Data is delivered on time and within budget Data is reliable and accessible to users when needed Data supports the project goals and use cases Data is discoverable	Number of users who have accessed the data Degree of data reusability in same or different domains Number of times the data is used in research, education, and commercial activities Number of novel projects, products, or breakthroughs enabled Number of contributors

Table 5. PSI Metrics for Processes

Progress	**Success**	**Impact**
% of systems that are monitored using a monitoring tool % of security incidents detected and responded to in a given time-frame % of maintenance tasks completed on time Mean Time To Detect (MTTD) security incidents Mean Time To Respond to security incidents % of systems that are patched and up to date % of users who have completed security and/or compliance training	Processes meet or exceed all requirements Processes are implemented on time and within budget Processes are effective in detecting, preventing, and responding to security incidents Processes support the organization's mission of providing reliable and secure CI resources to users User satisfaction with the new processes	Number of security incidents prevented Number of security incidents detected and addressed Downtime reduction Enhanced team productivity Cost savings Improvement in the quality of the products/research/projects

Table 5. PSI Metrics for People Related Projects

Progress	**Success**	**Impact**

Accepted manuscript

ACCEPTED MANUSCRIPT

Number of people who have completed trainings Number of people who have received certifications Number of instructors engaged Number of training modules or courses developed or enhanced Number of trainings or courses that were offered Number of training modules or courses that are digitally accessible Number of email lists or professional organizations or student groups contacted for announcing the availability of trainings or courses Number of activities for broadening participation	Training programs or courses meet or exceed all requirements Training programs or courses are delivered on time and within budget Training programs or courses are effective in increasing the knowledge and skills of participants The organization has achieved its goal of increasing workforce diversity and inclusiveness The organization has achieved its goal of increasing the diversity in leadership positions User satisfaction with training programs	Number of people who have been placed in new jobs or promoted as a result of trainings or courses Number of people who are employed in CI-related jobs as a result of trainings or courses Number of people from underrepresented groups who have been hired into CI-related jobs due to trainings or courses Number of people who are certified in CI-related skills Number of people who are using the CI skills learned to solve real-world problems Content follows W3C digital accessibility guidelines Increased skills of the instructors Increased public awareness of the technical advancements enabled

7. MICI - A Metrics Model for CI Projects

Metrics management within organizations is often highly evolutionary - the metrics captured, reported and analyzed are those which are necessary to satisfy operational management needs, financial management needs, and organizational performance management needs. Hence, there is a proliferation of point-to-point and *ad hoc* flows of metrics, and the secondary or derived metrics are cobbled together from the available metrics, with some duplication or overlap. Further, at some point, a comprehensive (and generally, complex) framework is retrofitted into the organic and evolutionary metrics, resulting in a tightly coupled framework that is fragile and difficult to manage.

It is desirable to have a comprehensive, yet lightweight, metrics framework that is flexible and easy to manage. In the case of CI projects, the metrics framework should also support standardization across CI projects. Therefore, to measure and manage the metrics associated with CI projects, we propose a model named MICI (Measuring the Impact of CyberInfrastructure). A high-level overview of the components of MICI is presented in Figure 3.

The metrics dimensions in the MICI Model are the PSI metrics that were described earlier. The metrics at each level are presented in Table 7. The processes associated with this model are described in the subsequent subsections.

Accepted manuscript

ACCEPTED MANUSCRIPT

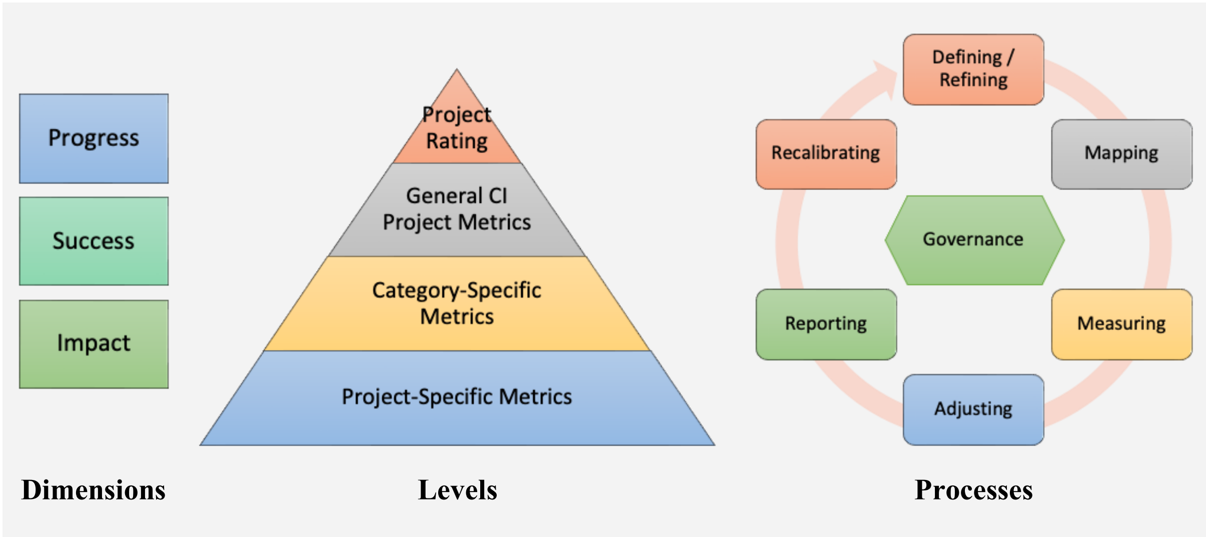


Figure 3. Overview of the MICI model

Table 7. Metrics for Different Pyramid Levels Shown in Figure 3

Level	**Metrics Type**	**Description**
Level 1	Project-Specific Metrics	Metrics unique to each CI project
Level 2	Category-Specific Metrics	Metrics common to a category of CI projects
Level 3	General Project Metrics	Top-level metrics, common to all projects
Level 4	Project Rating	A numeric rating derived from Level 3 metrics

7.1. Definition and Mapping of the Metrics

In order to fully define the metrics for a CI project in accordance with the MICI Model, the following steps should be followed:

1. Define the metrics necessary for measuring and managing the project, based only on the nature of the project, and without considering any frameworks, models, or standards. These will be the Level 1 metrics.
2. Reorganize the metrics defined into the following three dimensions: Progress, Success, and Impact. If necessary, add, remove, split or merge the metrics, until there is a meaningful way to measure all three dimensions.
3. Identify the category of the CI project and compare the Level 1 PSI metrics (that is, the metrics that are defined specifically for the CI project) with the Level 2 metrics (that is, the category-specific metrics). Create a mapping between the Level 1 and Level 2 metric such that the Level 2 metrics can be calculated if the Level 1 metrics are known. If a mapping is not easily apparent, consider adding additional L1 metrics, or perhaps selecting a different category for the project.

Accepted manuscript

4. Add Level 2 weights - between 0 and 1 - to the Level 2 metrics such that the sum of all Level 2 weights is 1.
5. Add Level 3 weights to the PSI dimensions based on the following rules:
   a. If the project is in progress, assign weights P=0.6, S=0.3 and I=0.1.
   b. If the project has ended, assign weights P=0, S=0.75 and I=0.25.
   c. Note that the Level 3 weights are dependent upon the state of the project at the time of evaluation.
   d. Within an organization or program, different rules for Level 3 weights may be adopted, reflecting the nature of projects therein, but if that is done, the weights will need to be reset/recalibrated before comparing with other organizations or programs.

As an example, a template illustrating the application of the MICI model for assessing the impact of a CI project in the software category is shown in Figure 4 below.

ABC Project - MICI Mapping

Project Name:	ABC Project
Other Project Info:	
MICI Category:	Software
Complete	N
Project Rating:	5.01

Project Metric Group	L1 (Project-Specific) Metric	L1 Metric Value	L1 Adjustment Factor	L1 Adjustment Rationale	L1 Adjusted Metric	L1 Metric Max	L1 Metric Type	L1 Score	L1 Weight	L2 (Category-Specific) Metric	L2 Score	L2 Weight	L3 (General CI) Metric	L3 Score	L3 Weight
Software Development	Services created	20	1.00		20.00	10.00		100.00%	1.00	% Modules developed	100.00%	0.33	Progress	69.40%	0.6
Software Development	Help pages/files created	20	0.50	This metric gets doubled because a rebranded copy of each page is maintained	10.00	5.00		100.00%	1.00	% Documentation Done	100.00%	0.33			
Data Analysis	Data points analyzed	5	1.00		5.00	50.00		10.00%	1.00	Reports prepared	10.00%	0.34			
Training & Outreach	Trainings Conducted	2	1.00		2.00	5.00		40.00%	1.00	Trainings Conducted	40.00%	0.50	Success	25.00%	0.3
Implementations	Use cases covered	1	1.00		1.00	10.00		10.00%	1.00	% Requirements Met	10.00%	0.50			
Publications	Number of publications	1	1.00		1.00	5.00		20.00%	1.00	Number of publications	20.00%	0.50	Impact	10.00%	0.1
Products	Domain-specific extensions	0	1.00		0.00	5.00		0.00%	1.00	Number of new products	0.00%	0.50			

Figure 4. A template to demonstrate the application of MICI model

In this example, the metrics mapping gets filled as follows:

1. In columns B, C and D, details of the L1 (project-specific) metrics are filled, which are based entirely on the nature and goals of the project, without any consideration of the MICI model. This includes the values for the L1 metrics at that time.
2. In column E, an adjustment factor is optionally specified (defaults to 1). This gives an opportunity to adjust the metric in case the metric is undesirably skewed for any reason. This adjustment should be done very sparingly, and a justification should be provided in column F, so that actual metrics do not get influenced by subjective opinions and motives.
3. In column G, the L1 Adjusted Metric is calculated automatically by multiplying the L1 Metric Value with the L1 Adjustment Factor. In column H, a maximum value for the L1 Metric is provided, so that no single metric can have an undesirably outsized impact on the overall scores.
4. In column I, the L1 Metric Type is specified. This helps in selecting the right formula to calculate the L1 Score. Currently, the following two types of metrics are used:
   a. Linear (default), in which a higher value is linearly correlated with a higher score. This is the default metric type, for which the column I entry can be left blank.
   b. Efficiency, denoted by "E" in column I, for which a lower value represents greater efficiency, and is inversely correlated with a higher score.
5. In column J, the L1 score is calculated automatically by the spreadsheet, using the following formulae:
   a. For a *Linear* metric type (column I is blank),
   $L1\ Score\ =\ Minimum\ (L1\ Adjusted\ Metric, L1\ Metric\ Max)\ /\ (L1\ Metric\ Max)$
   b. For an *Efficiency* metric type (column I contains "E"), the L1 Score is calculated using the formula for efficiency (Output/Input), but with the lower limit of input set to half of the

Accepted manuscript

'maximum':

$$L1\ Score\ =\ (L1\ Metric\ Max)\ /\ MAX\ (L1\ Adjusted\ Metric, (L1\ Metric\ Max)/2)$$

6. In column K, the L1 weight needs to be entered. This determines how much an L1 metric contributes towards an L2 metric. If the mapping of the L1 and L2 metrics is 1-to-1, the L1 weights will be exactly "1.00". Otherwise, some of them may be less than 1.
7. In column L, L2 (category-specific) metrics corresponding to the L1 metrics are entered. Based on columns J and K, the L2 Scores in column M are calculated by the spreadsheet, using the following formula:

$$L2\ Score\ =\ L1\ Score\ x\ L1\ Weight$$

8. In column N, the L2 weights are entered. These define the contribution of each L2 metric to the L3 metrics (general CI metrics - Progress, Success and Impact).
9. In column P, L3 scores for Progress, Success and Impact are calculated automatically by the spreadsheet, using the following formula:

$$L3\ Score_{(PSI\ Metric)} = \quad _{(PSI\ Metric)}\ (L2\ Score\ x\ L2\ Weight)$$

10. In column Q, weights to the three general metrics dimensions are given, based on the current stage of the project.
11. Finally, the Project Rating is calculated from columns P and Q, by taking the weighted average of the L3 scores, and mapping the result to a scale of 0 - 10. In the example shown in Figure 4:

$$Project\ Rating\ =\ (L3\ Score\ x\ L3\ Weight)\ x\ 10$$
$$= (\ (0.694\ x\ 0.6)\ +\ (0.25\ x\ 0.3)\ +\ (0.10\ x\ 0.1)\ )\ x\ 10$$
$$= 5.01$$

7.2. Capturing Metrics Data

Data for metrics can be gathered and aggregated in a database either manually or automatically depending upon the types of metrics chosen or defined for the CI projects and their related categories. As an example, scripts can be written for automatically gathering the data related to the utilization of hardware platforms and generating reports. Such data can be exported to CSV files which can then be exported to databases. When gathering data related to qualitative metrics, such as the number of participants of a training event who are actually applying the knowledge gained during training to solve real-world problems, one may need to administer surveys or interview those trainees. Certain metrics - such as the number of mentions of a project in the news or on social-media platforms - can be collected automatically via web-scraping tools.

The frequency at which the metrics should be collected must be clearly defined for the projects along with the methods and tools used for data collection. It should be noted that in the case of hardware and software systems, it can be difficult to accurately capture data for certain metrics after a long time has elapsed from the actual occurrence of certain events (such as, the number of customers using a software application at a given point in time). Hence, advance planning should be done for data collection for the relevant metrics and appropriate frequency of data collection or data sampling should be determined. For studying trends over a period of multiple years, it is important to keep copies of raw data in addition to saving the aggregations to ensure that there is no loss of information. For transparency, the data collected for metrics can be deposited in a central database such as the one supported by the Opuntia project [11]. Certain metrics such as the publicly available GitHub insights can be automatically gathered for a project cataloged in Opuntia if the PI provides the GitHub URL of their open-source projects. The number of times a software is used on a platform can be captured using tools like XALT [12]. The iTracker libraries developed in the Opuntia project can also be used for tracking the actual software use if the libraries are linked with the code at the time of code development and installation. These libraries track the number of times a software is used and if internet connectivity is available, can send the actual usage numbers to the Opuntia database.

Accepted manuscript

ACCEPTED MANUSCRIPT

The development of the feature for obtaining user consent before sending the data to a central repository is a part of future work for iTracker.

7.3. Adjusting and Reporting

While the automatic capturing and reporting of the metrics data is desirable, expert review and calibrations are also valuable for analyzing the standardized metrics across different projects and specifically within the context of a project. It is recommended to have adjustment factors and pre-defined weights, that help customize the metrics model to project-specific needs. The pre-defined weights and adjustment factors can be applied automatically, with an opportunity for manual adjustment. Ideally, the adjustments should be kept separate from the actual metrics data and any automatic adjustments made. Also, the adjustments made should be justified and approved by the PIs when reporting the metrics data to a central repository.

7.4. Recalibration

Weights and adjustment factors can be modified from time to time, to improve the usability of the metrics and reduce the need for manual adjustments. Care should be taken not to abuse the flexibility of the model by applying extremely large adjustment factors (such as 400% or more) that can dramatically change the results and diminish the reliability of the metrics. Manually adjusting the metrics data instead of automatically applying the adjustment factors, and constantly changing the adjustment factors themselves can result in unreliable metrics and trends. Hence, such activities should be governed and a process similar to the change request management process for software changes can be adopted.

7.5. Governance

The capture and use of metrics is generally not a fully automated process and requires continuous monitoring and adjustments. Further, the metrics themselves may need to be modified or recalibrated from time to time. To do that, an effective governance model is critical. For a typical academic or research organization, a governance model such as the one shown in Figure 5 is recommended. Note that a governance model for metrics that is presented here will primarily be useful only at an organization or program level, where a number or projects are being executed over a relatively long period of time. For a single project, setting up a new governance model is not necessary. Instead, the project is expected to be governed under a program or organization-level governance model. As with most governance frameworks, a few key elements are required and those are shown in Table 8.

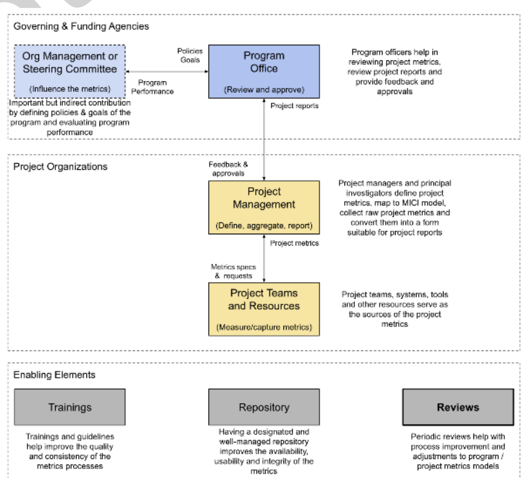


Figure 5. Overview of the governance structure

Table 8. Overview of key elements in the governance framework

ACCEPTED MANUSCRIPT

Governance Element	**Addresses the Key Concern**	**Recommendations**
Roles	*Who* will be involved in the capture, management and use of CI project metrics?	The following roles need to be fulfilled in the governance model presented ● Program Officer ● Project Manager ● Project Resource
Responsibilities	*What* will be the primary responsibilities of each role?	● Program Officer: Approves project metrics, reviews project reports and provides feedback and other approvals ● Project Manager: Defines project metrics, maps them to the MICI model, collects raw project metrics and converts them into a form suitable for project reports ● Project Resource: Any system, team member, organization or other source of project metrics
Repository	*Where* will the metrics be stored?	A secure and managed location for the metrics is highly recommended, so that availability, usability and integrity of the metrics is maintained. The repository should preferably be at the program or funding agency level, so that the complexity of numerous heterogeneous project-specific repositories is avoided.
Trainings and Reviews	*How* will the governance be affected?	Trainings: Project managers (including grant applicants) should be provided training or guidance on how to formulate project metrics, map them to category-level metrics, and how to aggregate and report the metrics. Reviews: Periodic assessments and adjustments of the metrics framework and its effectiveness for various projects.

This governance model for CI project metrics depicts the key participants from governing/funding agencies and from the organizations that are executing the project, along with their primary responsibilities. The

Accepted manuscript

ACCEPTED MANUSCRIPT

"Reviews" depicted are of special significance, as they provide an opportunity to modify or adjust the metrics being captured and how they are being captured and reported, based on the learnings from multiple projects. Good governance requires controlling unwarranted deviations from the strategies, as well as adjusting the strategies, when warranted. The reviews are critical to both these aspects of governance.

8. An Open-Infrastructure for Gathering Metrics of CI Projects and MICI

We have observed that the software and data products resulting from the CI projects may not have the sufficient level of visibility in the community and may not be easily findable, thereby restricting their potential of reuse in the community or getting cited by others. Therefore, a central repository is needed where the products resulting from CI projects can be cataloged along with any related code, data, test cases, and **metrics for assessment**. Towards meeting this need, we are building an open-infrastructure named as Opuntia [11], NSF award # 231420, for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products. Opuntia belongs to the "software" category of our CI project taxonomy. Opuntia will be beneficial for both the community and the funding agencies. While Opuntia can help in gathering and understanding the different types of metrics used by the different CI projects, it does not currently have a mechanism for using the metrics to derive the information on the impact created by the products/projects. Therefore, integrating a framework for assessing impacts of the CI projects with a platform like Opuntia will be useful towards creating an open and transparent data infrastructure as mentioned in Section 1 and is a part of our future work. A high-level overview of the design of the open-infrastructure for assessing the CI projects is shown in Figure 6.

In Figure 6, we show the different components of Opuntia, and from amongst these, the **catalog and iTracker** can help in automatically and manually gathering the publicly accessible data related to the metrics of interest. Opuntia is designed to collect data related to any user-defined metric and hence, the data related to both project-specific and standard metrics for the "software" category can be deposited in it.

The community and funding agencies can browse through the catalog of projects in Opuntia along with their metrics, and study trends as needed. Once the MICI model is integrated with Opuntia, we will be able to demonstrate the short-term and long-term impact of the different categories of CI projects through high-level interfaces.

Accepted manuscript

ACCEPTED MANUSCRIPT

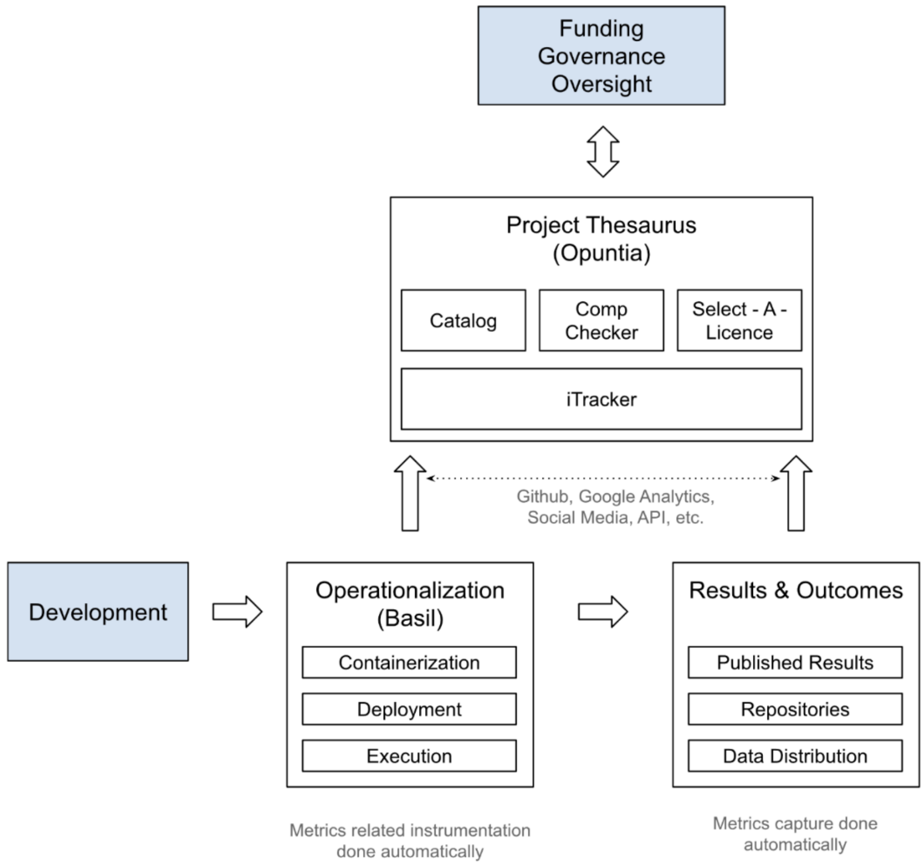


Figure 6. Design of the Open-Infrastructure for Assessing the CI Projects

8.1. Software Category Project as an Example

In Figure 6, we show a CI project named Basil [13], NSF award # 2314203, that belongs to the "software" category in our taxonomy. Basil supports the semi-automatic containerization of software and data products in a stand-alone mode and through a web-portal. We are using Basil here as a sample CI project that is contributing its metrics data in Opuntia. The project-specific metrics for assessing Basil are shown in Table 9. The "software" category-specific PSI metrics that are applicable to Basil are mentioned in Table 3. These are mapped, where applicable, to the project-specific metrics for Basil and this mapping is shown in Table 9 as well. Basil is a work in progress and is a three-year project that is currently in its second year. Hence, certain metrics like "ratings on surveys" are yet to be captured. Hence, TBD is used to represent the values of yet to be captured metrics. Currently, only GitHub related insights for the Basil project are shared through Opuntia but in future, all the other project-specific metrics and software-category specific metrics as mentioned above will be shared publicly.

Table 9. Basil Software - Mapping L1, L2, and L3 Metrics

Project-Specific Metrics (L1)	**Category-Specific Metrics (L2)** Software (SW) and People (WD) categories selected, category names and PSI dimensions mentioned	**General Metrics (L3)** PSI dimensions mentioned

Accepted Manuscript

ACCEPTED MANUSCRIPT

Number of unique users of Basil through the web-portal	Increase in number of adopters or users from diverse domains (**SW, I**) A reasonable size of user community exists (**SW, S**)	Acceptance/adoption by customers or users (**S**) Community engagement **(I)**
Number of Docker/Singularity images produced	Number of novel projects, products, or breakthroughs enabled (**SW, I**)	Domain impact (**I**)
Number of applications that are containerized	Software supports the targeted business functions (**SW, S**) Number of novel projects, products, or breakthroughs enabled (**SW, I**)	Outcomes - meeting project goals (**S**) Domain impact (**I**)
Number of science domains that are supported	Software supports the targeted business functions (**SW, S**) Number of adopters or users from diverse domains (**SW, I**) Number of novel projects, products, or breakthroughs enabled (**SW, I**) - **TBD**	Outcomes - meeting project goals (**S**) Domain impact (**I**)
% completion of the user-guide, course content, and video-demos	% completion of development of user guide, training content, or education and engagement activities for increasing software adoption (**SW, P**)	Milestones achieved (**P**)
Number of trainings/webinars conducted	Number of trainings or courses that were offered (**WD, P**) Increase in number of adopters or users from diverse domains (**SW, I**)	Milestones achieved (**P**) Outcomes - meeting project goals (**S**) Community engagement (**I**)
Ratings on surveys	User satisfaction with software features and functionality (**SW, S) - TBD**	Satisfaction - positive feedback from users and stakeholders (**S**) Community engagement (**I**)
Person months taken to develop the core features (in months)	Time taken to develop and deploy new software features (**SW, P)**	Milestones achieved (**P**)

Accepted manuscript

ACCEPTED MANUSCRIPT

Number of commits on GitHub	Number of code commits or changes or bug fixes in the code repository (**SW, P)**	Technological advancement **(I)**
Number of views of the YouTube videos for trainings and video-demonstrations	Increased public awareness of the technical advancements enabled (**WD, I**)	Media coverage **(I)** Community engagement (altmetrics) **(I)**
Number of social-media tweets, blogs, and news articles	Increased public awareness of the technical advancements enabled (**WD, I**)	Community engagement (altmetrics) **(I)**
Number of visitors of the web-portal	A reasonable size of user community exists (**SW, S**) Increase in number of adopters or users from diverse domains (**SW, I**)	Community engagement (altmetrics) **(I)**
Number of courses enhanced as a result of the project	Number of training modules or courses developed or enhanced (**WD, P**) Increase in number of adopters or users from diverse domains (**SW, I**)	Community engagement (**I**) Outcomes - meeting project goals (**S**)
Number of workshops for community engagement	Number of trainings or courses that were offered (**WD, P**) Increase in number of adopters or users from diverse domains (**SW, I**)	Community engagement (**I**) Outcomes - meeting project goals (**S**)

For the metrics defined in Table 9, we use the actual metric values and weights and derive an overall score for the Basil project using the "MICI Mapping" template for the project. As the Basil project is in its second year of development at the time of writing this paper and hence all the values for the metrics related to it are not yet available. An image of the filled template is shown in Figure 7 and the MICI scores for the Basil project are as follows:

- Progress: 97.38%
- Success: 25.30%
- Impact: 35.06%
- Overall project score (out of 10): 6.95

Accepted manuscript

ACCEPTED MANUSCRIPT

Basil - MICI Mapping			
Project Name:	Basil		
Other Project Info:			
MICI Category:	Software		
Complete	N		
Project Rating:	6.95	**Date:**	12/12/2023 (Year-2 of the project)

Project Metric Group	L1 (Project-Specific) Metric	L1 Metric Value	L1 Adjustment Factor	L1 Adjustment Rationale	L1 Adjusted Metric	L1 Metric Max	L1 Metric Type	L1 Score	L1 Weight	L2 (Category-Specific) Metric	L2 Score	L2 Weight	L3 (General CI) Metric	L3 Score	L3 Weight
Community engagement	% completion of the user-guide, course content, and video-demos	100	1.00		100.00	100.00		100.00%	1.00	% completion of development of user guide, training content, or education and engagement activities for increasing software adoption	100.00%	0.25	Progress	97.38%	0.6
Community engagement	Number of trainings / webinars conducted	2	1.00		2.00	3.00		66.67%	0.50	Number of trainings or courses that were offered	33.33%	0.10			
Technological advancement	Person months taken to develop the core features (in months) thus far	22	1.00	Lower value is better as it indicates the work completion in lesser time than originally estimated and hence adjusting the formula to MAX (L1 Adjusted Metric, L1 Metric Max)/L1 Metric Max	22.00	33.50	E	152.27%	1.00	Time taken to develop and deploy new software features	152.27%	0.17			
Technological advancement	Number of commits on GitHub	571	1.00	Listing minimum number of commits instead of maximumfor "L1 Metric Max"	571.00	25.00		100.00%	1.00	Number of code commits or changes or bug fixes in the code repository	100.00%	0.17			
Community engagement	Number of courses enhanced as a result of the project	1	1.00		1.00	1.00		100.00%	1.00	Number of training modules or courses developed or enhanced	100.00%	0.22			
Community engagement	Number of workshops for community engagement	1	1.00		1.00	1.00		100.00%	0.50	Number of trainings or courses that were offered	50.00%	0.10			
Community engagement	Number of unique users of Basil through the web-portal	33	0.75	Adjusted, as auto-reported user count contains 'test' users and also considering "L1 Metric Minimum" value instead of "L1 metric Max"	24.75	50.00		49.50%	0.50	A reasonable size of user community exists	24.75%	0.20	Success	25.30%	0.3
Domain impact	Number of applications that are containerized	13	1.00		13.00	50.00		26.00%	0.50	Software supports the targeted business functions	13.00%	0.25			
Domain impact	Number of science domains that are supported	6	1.00		6.00	10.00		60.00%	0.50	Software supports the targeted business functions	30.00%	0.25			
Community engagement	Ratings on surveys	0	1.00	Surveys are yet to be conducted	0.00	5.00		0.00%	1.00	User satisfaction with software features and functionality	0.00%	0.10			
Community engagement (altmetrics)	Number of visitors of the web-portal	480	1.00		480.00	500.00		96.00%	0.50	A reasonable size of user community exists	48.00%	0.20			
Community engagement	Number of unique users of Basil through the web-portal	33	1.00		33.00	50.00		66.00%	0.50	Number of adopters or users from diverse domains	33.00%	0.16	Impact	35.06%	0.1
Domain impact	Number of Docker / Singularity images produced	52	1.00		52.00	50.00		100.00%	0.33	Number of novel projects, products, or breakthroughs enabled	33.00%	0.12			
Domain impact	Number of applications that are containerized	13	1.00	The data related to the "novel" aspect of the L2 metric is yet to be captured. The number of test cases that are containerized is 22 and the number of applications containerized is 13.	13.00	20.00		65.00%	0.33	Number of novel projects, products, or breakthroughs enabled	21.45%	0.12			
Domain impact	Number of science domains that are supported	6	1.00		6.00	10.00		60.00%	0.50	Number of adopters or users from diverse domains	30.00%	0.16			
Domain impact	Number of science domains that are supported	6	1.00		6.00	10.00		60.00%	0.34	Number of novel projects, products, or breakthroughs enabled	20.40%	0.12			
Community engagement (altmetrics)	Number of views of the YouTube videos for trainings and video-demonstrations	233	1.00		233.00	100.00		100.00%	0.50	Increased public awareness of the technical advancements enabled	50.00%	0.16			
Community engagement (altmetrics)	Number of social-media tweets, blogs, and news articles	14	1.00	7 tweets and 2 blogs are the minimum thresholds for this metric to be successful. 11 tweets and 3 LinkedIn.com posts (akin to blogs) have been made.	14.00	9.00		100.00%	0.50	Increased public awareness of the technical advancements enabled	50.00%	0.16			

Figure 7. MICI Mapping template for the Basil Project

8.2 . People Category Project as an Example

Let us consider a hypothetical CI project that belongs to the "people" category of our taxonomy. This project is meant for broadening participation in a High Performance Computing (HPC) summer school that is organized every year at an institution. The problem statement for this project is that, historically, there

ACCEPTED MANUSCRIPT

has been a lack of diversity at the summer school, and therefore, the next summer school should be made inclusive and diverse by creating participation opportunities for individuals from underrepresented groups and Minority-Serving Institutions (MSIs) and financially supporting their participation. In this example, the stakeholders for the summer school include the organizing committee, head of the hosting institution, instructors, and the participants. Some of the **processes or activities** for achieving the project goals include: (1) training the instructors to create an inclusive teaching environment and content/curricula, (2) reaching out to the student affinity groups such as ACM-W [14] to advertise the availability of the participant support, and (3) making the meeting rooms accessible. The expected **output or success** of this project can include a summer school that is inclusive and has a diverse group of participants. If this project is successful, as its **impact or outcome**, it would have reduced the barriers to creating equal learning opportunities. A post-event survey can help in gathering the quantitative and qualitative metrics for measuring the outcome or the impact created. Some of the **metrics for assessing the PSI** of this project are shown in Table 10.

Table 10. People Project- Mapping L1, L2, and L3 Metrics

Project-Specific Metrics (L1)	**Category-Specific Metrics (L2)** People (WD) category selected**,** category names and PSI dimensions mentioned	**General Metrics (L3)** PSI dimensions mentioned
% of instructors who have completed the trainings on developing inclusive curricula and learning environments	Number of people who have completed trainings (**WD, P**) Number of instructors engaged (**WD, P**) Content follows W3C digital accessibility guidelines (**WD, I**)	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**
Number of student affinity groups that were contacted to advertise the availability of participant support	Number of email lists or professional organizations or student groups contacted for announcing the availability of trainings or courses (**WD, P**)	Community engagement **(I)**
Number of participants from underrepresented groups that were selected to participate in the summer school	Number of activities for broadening participation (**WD, I)**	Community engagement **(I)**
The number of updates or modifications made to the policies to make the meeting rooms accessible	Number of training modules or courses developed or enhanced (**WD, P**)	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**

Accepted manuscript

ACCEPTED MANUSCRIPT

Number of stakeholder meetings that were held to ensure that the project activities themselves are designed in an inclusive manner	Number of activities for broadening participation **(WD, P)**	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**
% increase in the number of participants from the underrepresented groups in the summer school as compared to the previous years	The organization has achieved its goal of increasing workforce diversity and inclusiveness **(WD, S)**	Outcomes - meeting project goals **(S)**
% increase in the number of accessible meeting rooms at the institution	Number of activities for broadening participation **(WD, P)**	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**
% increase in the number of instructors who presented inclusive content	Increased skills of the instructors **(WD, I)**	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**
Are any incidences reported on microaggression?	The organization has achieved its goal of increasing workforce diversity and inclusiveness **(WD, S)**	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities **(I)**
Number of participants from underrepresented groups who reported feeling included	The organization has achieved its goal of increasing workforce diversity and inclusiveness **(WD, S)**	Satisfaction - positive feedback from users and stakeholders **(S)**
Positive intention of the participants from the underrepresented groups to apply the knowledge gained at the summer school in research and scholarly activities	Training programs are effective in increasing the knowledge and skills of participants **(WD, S)** Number of people who are using the CI skills learned to solve real-world problem **(WD, I)** The organization has achieved its goal of increasing workforce	Satisfaction - positive feedback from users and stakeholders **(S)**

Accepted manuscript

	diversity and inclusiveness (**WD, S**)	
Implementation of a new policy for developing inclusive curricula (*e.g.*, following W3C accessibility guidelines) at the host institution	Number of training modules or courses that are digitally accessible (**WD, P**) Content follows W3C digital accessibility guidelines (**WD, I**)	Institutional impact - such as, enhancing resources, reputation, skills, and capabilities (**I**)

For the metrics defined in Table 10, we assume metric values and weights and derive an overall score for the project using the "MICI Mapping" template for the project. An image of the filled template is shown in Figure 8 and the MICI scores for the HPC Summer School project are as follows:

- Progress: N/A (project complete)
- Success: 70.00%
- Impact: 89.58%
- Overall project score (out of 10): 7.49

9. Required Calibrations, Limitations, and Future Work

Each project has a certain budget, scope of work, and a time-line for completion. It is important to calibrate the impact score calculated by the MICI model according to the total cost, effort, and time spent on the project after the project is over to determine the impact per unit of cost/effort/time. Additionally, as a project progresses towards completion, its impact score will change. The assessment score per year should be tracked to establish baselines and find trends indicating progress, success, and impact over the project lifecycle and beyond. Such data will help in determining the long-term impact of the projects.

The MICI Model cannot ensure that metrics for different projects have been equitably designed. It will be necessary for the sponsors and governors of the projects to validate the proposed metrics and allow only those that are appropriate for the project. Otherwise, the metrics can be misleading when compared to other projects. However, this is not a problem that is introduced by the MICI Model. Rather, it is something which may already be happening, and will become more obvious when a common yardstick like the MICI Model starts getting used. In fact, it is one of the main motivating factors behind developing this model: We believe that an unreasonably high or unreasonably low bar for performance/success/impact of a project leads to sub-optimal utilization (or wastage) of precious funding and intellectual capital. This is because an exceedingly easy target generally does not inspire groundbreaking outcomes, whereas an impractically difficult target may be discouraging or unattainable. Therefore, we believe there is immense value in having a common metrics model that can be used across projects, while still allowing the flexibility to customize it for specific projects. We believe it will be a step away from a 'project-by-project' evaluation mindset, towards relative performance mindset, which helps us achieve two very important goals: (1) formulating project metrics that maximize the advancement/outcomes from the project; and (2) better recognizing and rewarding performance and impact, so that funding can be directed where there will be maximum ROI.

Accepted manuscript

ACCEPTED MANUSCRIPT

As a part of our future work, we will develop complete recipes for gathering different types of metrics for CI projects and discuss a list of denominators that can be used to calibrate the projects across funding programs and institutions. We will also enhance Opuntia - the open data infrastructure described in this paper - to support automatic collection of the altmetrics related to the CI projects. We will also share our work on, (1) evolving the MICI Model, (2) providing a simple, spreadsheet-based tool for manual capture, storage, audit and self-assessment of CI project metrics, (3) enhancing and integrating the Opuntia and Basil software to understand and automate the metrics management process, covering metrics definition, capture, storage, use, reporting, analysis and governance subprocesses, and (4) inviting collaborations and organizing events (such as, the Metrics2023 Conference [15]), to drive community contributions, refinements, support, and adoption of the MICI model. A special working group can potentially be formed to prepare an exhaustive list of metrics across the different categories of the CI projects.

HPC Summer School - MICI Mapping

Project Name:	**HPC Summer School**		
Other Project Info:			
MICI Category:	Software		
Complete	Y		
Project Rating:	**7.49**	**Date:**	December 12, 2023

Project Metric Group	L1 (Project-Specific) Metric	L1 Metric Value	L1 Adjustment Factor	L1 Adjustment Rationale	L1 Adjusted Metric	L1 Metric Max	L1 Metric Type	L1 Score	L1 Weight	L2 (Category-Specific) Metric	L2 Score	L2 Weight	L3 (General CI) Metric	L3 Score	L3 Weight
Institutional impact	Number of instructors who have completed the trainings on developing inclusive curricula and learning environments	8	1.00		8.00	8.00		100.00%	1.00	Number of people who have completed trainings	100.00%	0.14	Progress	88.46%	0
Institutional impact	Number of instructors who have completed the trainings on developing inclusive curricula and learning environments	8	1.00		8.00	8.00		100.00%	1.00	Number of instructors engaged	100.00%	0.14			
Community engagement	Number of student affinity groups that were contacted to advertise the availability of participant support	11	1.00		11.00	15.00		73.33%	1.00	Number of email lists or professional organizations or student groups contacted for announcing the availability of trainings or courses	73.33%	0.14			
Institutional impact	The number of updates or modifications made to the policies to make the meeting rooms accessible	1	1.00		1.00	1.00		100.00%	1.00	Number of training modules or courses developed or enhanced	100.00%	0.19			
Institutional impact	Number of stakeholder meetings that were held to ensure that the project activities themselves are designed in an inclusive manner	5	1.00		5.00	6.00		83.33%	0.50	Number of activities for broadening participation	41.67%	0.07			
Institutional impact	% increase in the number of accessible meeting rooms at the institution	50	1.00		50.00	50.00		100.00%	0.50	Number of activities for broadening participation	50.00%	0.07			
Institutional impact	Implementation of a new policy for developing inclusive curricula (e.g., following W3C accessibility guidelines) at the host institution	2	1.00		2.00	1.00		100.00%	1.00	Number of training modules or courses that are digitally accessible	100.00%	0.24			
Outcomes - meeting project goals	% increase in the number of participants from the underrepresented groups in the summer school as compared to the previous years	20	1.00		20.00	20.00		100.00%	0.25	The organization has achieved its goal of increasing workforce diversity and inclusiveness	25.00%	0.10	Success	70.00%	0.75
Institutional impact	Are any incidences reported on microaggression? No incidences reported or number of incidences reported that are less than the maximum value defined means success. However, if 0 is selected as the metric value, it can lead to divide by 0 errors or 0% progress in the general formula. Hence, no incidence reported is represented as 1 here.	1	1.00		1.00	1.00		100.00%	0.25	The organization has achieved its goal of increasing workforce diversity and inclusiveness	25.00%	0.10			
Satisfaction - positive feedback from users and stakeholders	Number of participants from underrepresented groups who reported feeling included	17	1.00		17.00	15.00		100.00%	0.25	The organization has achieved its goal of increasing workforce diversity and inclusiveness	25.00%	0.10			
Satisfaction - positive feedback from users and stakeholders	Positive intention of the participants from the underrepresented groups to apply the knowledge gained at the summer school in research and scholarly activities	47	1.00		47.00	40.00		100.00%	1.00	Training programs are effective in increasing the knowledge and skills of participants	100.00%	0.60			
Satisfaction - positive feedback from users and stakeholders	Positive intention of the participants from the underrepresented groups to apply the knowledge gained at the summer school in research and scholarly activities	47	1.00		47.00	40.00		100.00%	0.25	The organization has achieved its goal of increasing workforce diversity and inclusiveness	25.00%	0.10			
Institutional impact	Number of instructors who have completed the trainings on developing inclusive curricula and learning environments	8	1.00		8.00	8.00		100.00%	0.50	Content follows W3C digital accessibility guidelines	50.00%	0.10	Impact	89.58%	0.25
Community engagement	Number of participants from underrepresented groups that were selected to participate in the summer school	59	1.00		59.00	60.00		98.33%	1.00	Number of activities for broadening participation	98.33%	0.25			
Institutional impact	% increase in the number of instructors who presented inclusive content	100	1.00		100.00	100.00		100.00%	1.00	Increased skills of the instructors	100.00%	0.25			
Satisfaction - positive feedback from users and stakeholders	Positive intention of the participants from the underrepresented groups to apply the knowledge gained at the summer school in research and scholarly activities	47	1.00		47.00	40.00		100.00%	1.00	Number of people who are using the CI skills learned to solve real-world problem	100.00%	0.30			
Institutional impact	Implementation of a new policy for developing inclusive curricula (e.g., following W3C accessibility guidelines) at the host institution	2	1.00		2.00	1.00		100.00%	0.50	Content follows W3C digital accessibility guidelines	50.00%	0.10			

Figure 8. MICI Mapping template for the HPC Summer School Project

10. Conclusion

CI project metrics are difficult to capture and represent in a standardized manner, due to factors such as the diversity and uniqueness of the projects. However, having a standardized metrics model for CI projects can be useful for developing an effective strategy for the administration of CI projects and programs, including definition of the projects, measuring progress of the projects, evaluating their success, and

ACCEPTED MANUSCRIPT

maximizing their impact. Additionally, funding decisions and governance of the projects can be better informed by these metrics. In this paper, we introduced the **CI Project Taxonomy** developed as part of the Opuntia project, discussed the open-infrastructure of Opuntia, and also discussed how Opuntia can be used for tracking, gathering, and aggregating the user-defined metrics of products released and deployed on different platforms and computing environments.

We also introduced the MICI Model that provides a *flexible but standardized* set of metrics for CI projects, enabling CI projects to quickly adopt a set of metrics applicable to their project type, customize the metrics further if necessary, and map them to a single, unified framework for measuring the success of the CI projects. Funding agencies such as NSF can apply the desired parameters to the MICI Model and analyze the data derived from it to support their decisions and operations. Concrete implementation of the MICI model, the guidelines for its adoption, and its potential integration with Opuntia are also discussed in this paper. The key points related to the implementation of the MICI model are that: (1) project-specific metrics such as hardware/software components, user counts and use cases (pilot adoptions) should be well-defined at the beginning of the project, (2) the project-specific metrics should be fully *mapped* to the MICI Model metrics, and (3) as the project progresses, it may become obvious that the project-specific metrics need to be revised or recalibrated and *remapped* to the MICI Model. The remapping of metrics demonstrates how the insights gained during project execution may result in adjustments to the path/approach, without compromising the original goals.

We should also note here that measuring the impact of the CI through "responsible metrics" [10] is important - metrics have the power to influence decision-making processes and assess impacts. They should not be just numbers or quantities for project management but should be robust enough to assess the quality and impacts of the outcomes of the project. As a community, we should ponder whether the metrics such as the number of downloads of software, journal impact factors, h-index, and number of impressions to social media posts are good measures for conveying the impact of the investments or projects. The outcomes of such community discussions should inform the CI projects and funding programs in future.

11. Statements and Declarations

Source of funding: We are grateful to NSF for award # 2314202 that has enabled us to develop the Opuntia project. This project has motivated us to think deeply about the metrics of the CI projects and identify gaps and opportunities.

Ethical approval: The work presented in this paper does not involve human subjects or any type of Institutional Review Board (IRB) approval.

Competing or financial interests: The authors are not employed by any organization that may gain or lose financially through publication of this manuscript.

References

1. NSF CSSI Program, website last accessed on December 1, 2023: https://new.nsf.gov/funding/opportunities/cyberinfrastructure-sustained-scientific
2. NSF SI2 Program, website last accessed on December 1, 2023: https://www.nsf.gov/pubs/2016/nsf16532/nsf16532.htm
3. NSF DIBBS Program website, website last accessed on December 1, 2023: https://www.nsf.gov/pubs/2017/nsf17500/nsf17500.htm
4. DORA, website last accessed on December 1, 2023: https://sfdora.org/
5. U.K. Research Excellence Framework, website last accessed on December 1, 2023: https://www.ref.ac.uk

Accepted manuscript

ACCEPTED MANUSCRIPT

6. Guidance on REF 2021 results, website last accessed on December 1, 2023: https://ref.ac.uk/guidance-on-results/guidance-on-ref-2021-results/
7. Snowball Metrics Recipe Book Standardised research metrics – by the sector, for the sector, website last accessed on December 1, 2023: https://snowballmetrics.com/wp-content/uploads/2022/07/0211-Snowball-Metrics-Recipe-Book-v7-LO-1.pdf
8. Julia Lane, stefano Bertuzzi, The STAR METRICS Project: Current and Future Uses for S&E Workforce Data, website last accessed on December 1, 2023: https://www.nsf.gov/sbe/sosp/workforce/lane.pdf
9. COARA, website last accessed on December 1, 2023: https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf
10. Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. DOI: 10.13140/RG.2.1.4929.1363
11. Opuntia, website last accessed on December 1, 2023: https://opuntia.online
12. Agrawal, K., Fahey, M. R., McLay, R., & James, D. (2014). User Environment Tracking and Problem Detection with XALT. In Proceedings of the First International Workshop on HPC User Support Tools (pp. 32–40). Piscataway, NJ, USA: IEEE Press. http://doi.org/10.1109/HUST.2014.6
13. Basil, website last accessed on December 1, 2023: https://icompute.us
14. ACM-W, website last accessed on December 1, 2023: https://women.acm.org/
15. Metrics2023 conference website, last accessed on April 9, 2024: https://sites.google.com/view/metrics2023

Accepted manuscript

ACCEPTED MANUSCRIPT

Appendix A: A Taxonomy for CI Projects

As part of the Opuntia project (NSF award # 2314202), the following taxonomy of CI projects has been developed, to serve as a starting point for defining an appropriate set of metrics for a given project.

1. Hardware or Instruments

- 1.1. Computing
  - 1.1.1. Cloud Computing
  - 1.1.2. Distributed Computing (Web Servers)
  - 1.1.3. Data-Intensive Computing
  - 1.1.4. High Performance Computing (HPC)
  - 1.1.5. IoT and Edge Computing
  - 1.1.6. Quantum Computing
  - 1.1.7. Volunteer Computing
  - 1.1.8. High-Throughput Computing
  - 1.1.9. Reconfigurable Computing (FPGAs)
- 1.2. Data Storage
- 1.3. Network and Communication Equipment
- 1.4. Remote Sensing

2. Software

- 2.1. Application Software
- 2.2. CI/CD and Process Automation
- 2.3. Cloud Computing Tools, Frameworks, and Environments
- 2.4. Code Optimization and Modernization Tool
- 2.5. Code Compression Tools
- 2.6. Cybersecurity
- 2.7. Content Management Tools
- 2.8. CRM tools
- 2.9. Data Management
  - 2.9.1. Data Archival and Preservation

Accepted manuscript

ACCEPTED MANUSCRIPT

2.9.2. Data Protection and Recovery

2.9.3. Data Privacy

2.9.4. Data Warehousing and Data Lake Platforms

2.9.5. Database and Data Processing

2.9.5.1. Distributed Databases

2.9.6. Semantic Web

2.10. Decision-Support System, Expert System, Knowledgebase

2.11. Embedded Software

2.12. Fault-tolerance

2.13. Filesystems and Parallel I/O

2.14. Generative programming tools and frameworks

2.15. High-level interfaces, libraries, compilers, and runtime systems for parallel programming

2.16. HPC Science Gateways (HPC in the Cloud)

2.17. Large-scale HPC applications (tuning, optimization, and implementation on HPC resources)

2.18. Image and Video Processing

2.19. Learning Management Systems

2.20. Measurement and Monitoring

2.21. Mobile Applications

2.22. Networking

2.23. Pattern Recognition

2.24. Programming Languages, Programming Environments, and Runtime Systems

2.25. Quantum Computing toolkits

2.26. Simulation Platforms

2.27. Software Configuration Management

2.28. Software Libraries

2.29. Software Verification and Validation

2.30. Supporting Software and Middleware for HPC environments

2.31. System and Network Management Software

2.32. Tools and techniques for Code Modernization

Accepted manuscript

ACCEPTED MANUSCRIPT

2.33. Tools and techniques for Memory and Power Optimization

2.34. Tools for Profiling, Debugging, and Parallelizing Applications

2.35. Tools for Supporting Collaborative and Virtual Environments, Virtual Assistance

2.36. Tools for Supporting Volunteer Computing and High-Throughput Computing

2.37. Tools and Environments for Scientific Visualization

2.38. Virtualization and Containerization Software

2.39. Web Portals, Web Services, Middleware, and Web-accessible Products

2.40. Workflow Management

3. Data

3.1. Structured Databases

3.2. Unstructured Data Collections

3.3. Semi-Structured Data

4. Processes

4.1. Software Engineering

4.2. Cybersecurity

4.3. Project Management

4.4. Automation

5. People

5.1. Broadening Participation

5.2. Management of Resources

5.3. Research Facilitation Services

5.4. Workforce Development

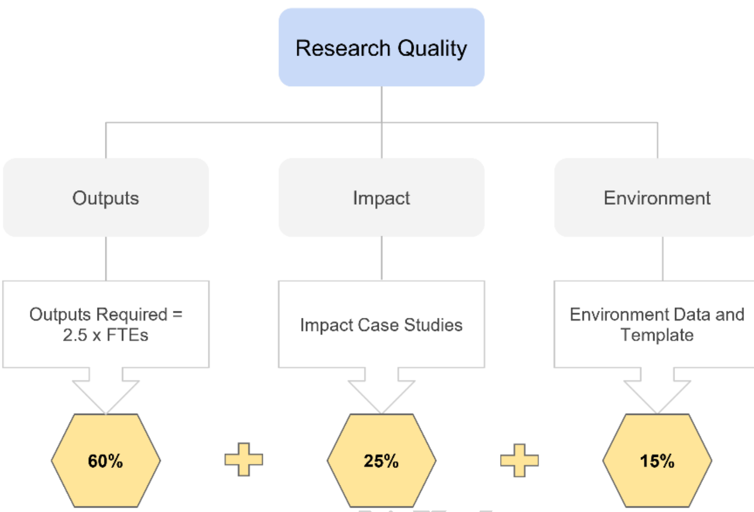
5.4.1. Training

5.4.2. Education

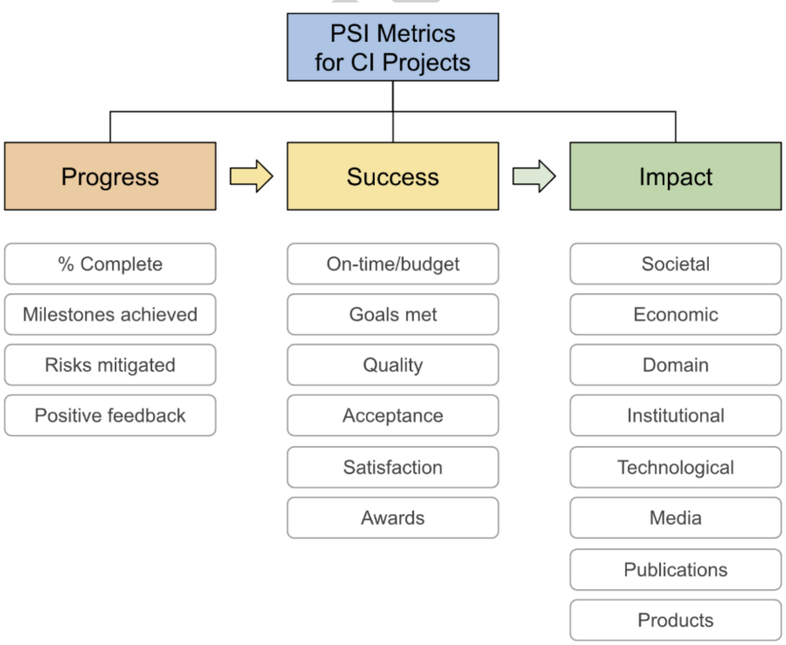
6. Combination of the above

Accepted manuscript

7. Other

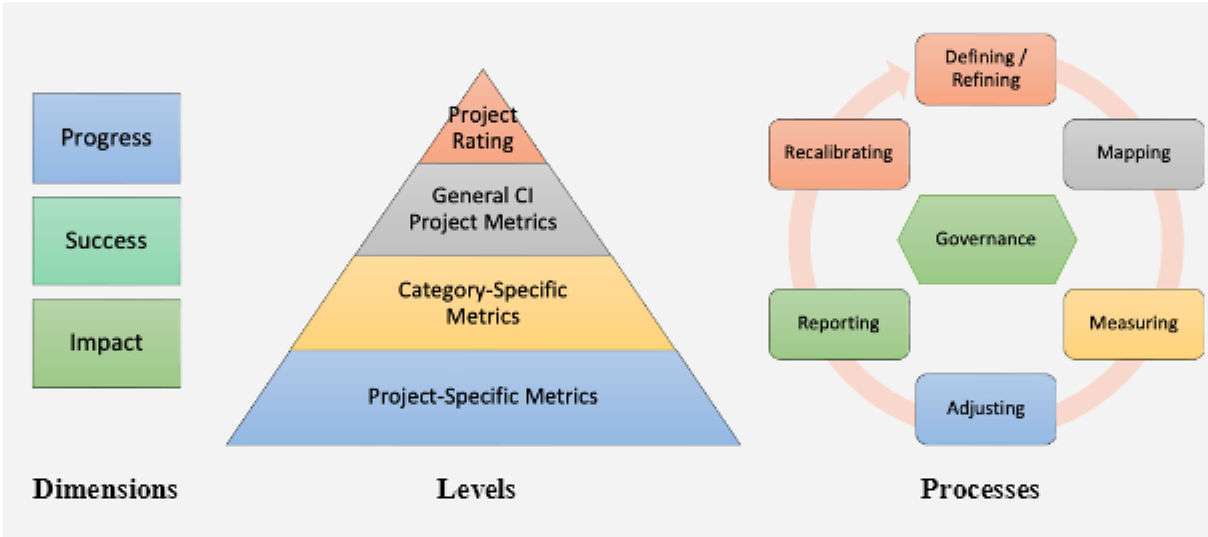


Research Quality
Outputs
Impact
Environment
Outputs Required = 2.5 x FTEs
Impact Case Studies
Environment Data and Template
60%
25%
15%



PSI Metrics for CI Projects
Progress
Success
Impact
% Complete
Milestones achieved
Risks mitigated
Positive feedback
On-time/budget
Goals met
Quality
Acceptance
Satisfaction
Awards
Societal
Economic
Domain
Institutional
Technological
Media
Publications
Products

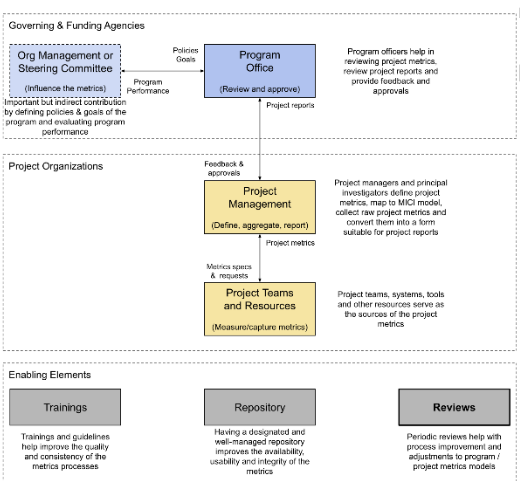
ACCEPTED MANUSCRIPT



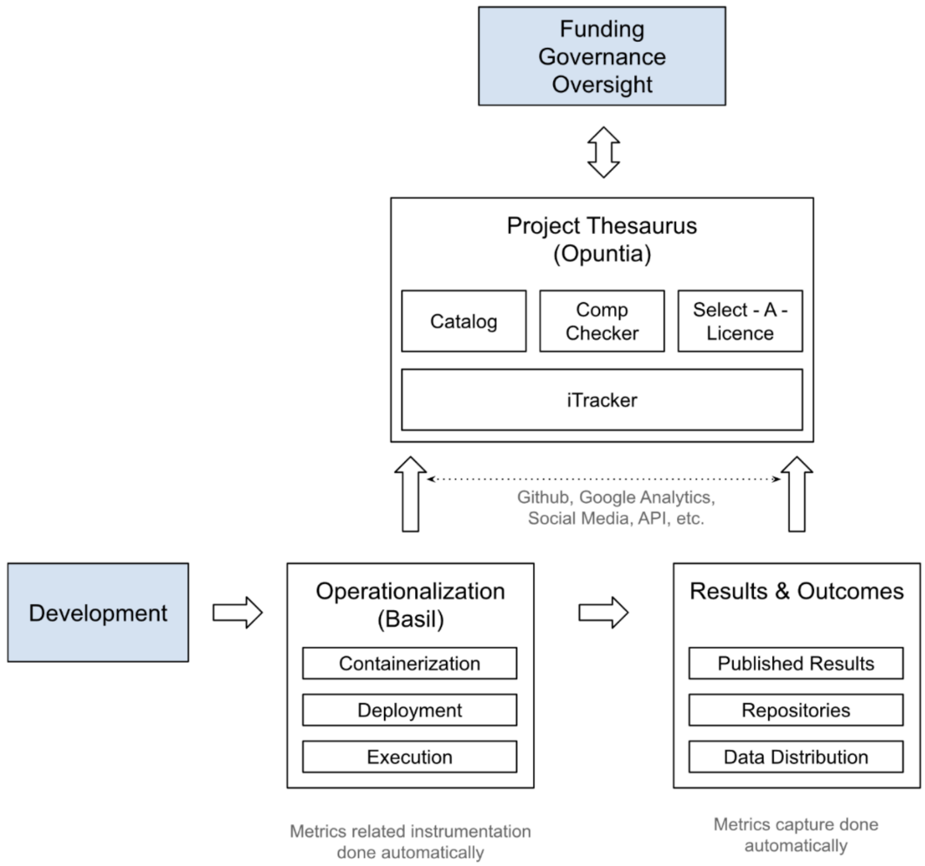
ABC Project - MICI Mapping

Project Name:	**ABC Project**
Other Project Info:	
MICI Category:	Software
Complete	N
Project Rating:	**5.01**

Project Metric Group	L1 (Project-Specific) Metric	L1 Metric Value	L1 Adjustment Factor	L1 Adjustment Rationale	L1 Adjusted Metric	L1 Metric Max	L1 Metric Type	L1 Score	L1 Weight	L2 (Category-Specific) Metric	L2 Score	L2 Weight	L3 (General CI) Metric	L3 Score	L3 Weight
Software Development	Services created	20	1.00		20.00	10.00		100.00%	1.00	% Modules developed	100.00%	0.33			
Software Development	Help pages/files created	20	0.50	This metric gets doubled because a rebranded copy of each page is maintained	10.00	5.00		100.00%	1.00	% Documentation Done	100.00%	0.33	Progress	69.40%	0.6
Data Analysis	Data points analyzed	5	1.00		5.00	50.00		10.00%	1.00	Reports prepared	10.00%	0.34			
Training & Outreach	Trainings Conducted	2	1.00		2.00	5.00		40.00%	1.00	Trainings Conducted	40.00%	0.50	Success	25.00%	0.3
Implementations	Use cases covered	1	1.00		1.00	10.00		10.00%	1.00	% Requirements Met	10.00%	0.50			
Publications	Number of publications	1	1.00		1.00	5.00		20.00%	1.00	Number of publications	20.00%	0.50	Impact	10.00%	0.1
Products	Domain-specific extensions	0	1.00		0.00	5.00		0.00%	1.00	Number of new products	0.00%	0.50			



ACCEPTED MANUSCRIPT



Accepted Manuscript

ACCEPTED MANUSCRIPT

Project Name:	**Basil**		
Other Project Info:			
MICI Category:	Software		
Complete	N		
Project Rating:	**6.95**	**Date:**	12/12/2023 (Year-2 of the project

Project Metric Group	L1 (Project-Specific) Metric	L1 Metric Value	L1 Adjustment Factor	L1 Adjustment Rationale	L1 Adjust Metri
Community engagement	% completion of the user-guide, course content, and video-demos	100	1.00		100.0
Community engagement	Number of trainings / webinars conducted	2	1.00		2.00
Technological advancement	Person months taken to develop the core features (in months) thus far	22	1.00	Lower value is better as it indicates the work completion in lesser time than originally estimated and hence adjusting the formula to MAX (L1 Adjusted Metric, L1 Metric Max)/L1 Metric Max	22.0
Technological advancement	Number of commits on GitHub	571	1.00	Listing minimum number of commits instead of maximumfor "L1 Metric Max"	571.0
Community engagement	Number of courses enhanced as a result of the project	1	1.00		1.00
Community engagement	Number of workshops for community engagement	1	1.00		1.00
Community engagement	Number of unique users of Basil through the web-portal	33	0.75	Adjusted, as auto-reported user count contains 'test' users and also considering "L1 Metric Minimum" value instead of "L1 metric Max"	24.7
Domain impact	Number of applications that are containerized	13	1.00		13.0
Domain impact	Number of science domains that are supported	6	1.00		6.00

ACCEPTED MANUSCRIPT

Project Name:	**HPC Summer School**
Other Project Info:	
MICI Category:	Software
Complete	Y
Project Rating:	**7.49**

Project Metric Group	**L1 (Project-Specific) Metric**	**L1 Metric Value**	**Adju Fa**
Institutional impact	Number of instructors who have completed the trainings on developing inclusive curricula and learning environments	8	
Institutional impact	Number of instructors who have completed the trainings on developing inclusive curricula and learning environments	8	
Community engagement	Number of student affinity groups that were contacted to advertise the availability of participant support	11	
Institutional impact	The number of updates or modifications made to the policies to make the meeting rooms accessible	1	
Institutional impact	Number of stakeholder meetings that were held to ensure that the project activities themselves are designed in an inclusive manner	5	
Institutional impact	% increase in the number of accessible meeting rooms at the institution	50	
Institutional impact	Implementation of a new policy for developing inclusive curricula (e.g., following W3C accessibility guidelines) at the host institution	2	
Outcomes - meeting project goals	% increase in the number of participants from the underrepresented groups in the summer school as compared to the previous years	20	
Institutional impact	Are any incidences reported on microaggression? No incidences reported or number of incidences reported that are less than the maximum value defined means success. However, if 0 is selected as the metric value, it can lead to divide by 0 errors or 0%	1	